The effect of high- and low-frequency previews and sentential fit on word skipping during reading

Bernhard Angele[1], Abby Laishley[1], Keith Rayner[2], and Simon P. Liversedge[3]

[1]Psychology Research Centre, Faculty of Science and Technology, Bournemouth University

[2]Department of Psychology, University of California San Diego

[3]School of Psychology, University of Southampton

Correspondence to:

Bernhard Angele

Psychology Research Centre

Faculty of Science and Technology, Bournemouth University

Talbot Campus

Fern Barrow, Poole BH12 5BB, United Kingdom

Email: bangele@bournemouth.ac.uk

Running Head: Skipping of short words

Abstract

In a previous gaze-contingent boundary experiment, Angele and Rayner (2012) found that readers are likely to skip a word that appears to be the definite article *the* even when syntactic constraints do not allow for articles to occur in that position. In the present study, we investigated whether the word frequency of the preview of a three-letter target word influences a reader's decision to fixate or skip that word. We found that the word frequency rather than the felicitousness (syntactic fit) of the preview affected how often the upcoming word was skipped. These results indicate that visual information about the upcoming word trumps information from the sentence context when it comes to making a skipping decision. Skipping parafoveal instances of *the* therefore may simply be an extreme case of skipping high-frequency words.

In order to allocate their gaze efficiently, readers have to anticipate how much information they will need about upcoming (that is, not yet fixated) words. There are two possible sources for information about upcoming words: First, readers can use parafoveal vision to preprocess words before they are fixated (McConkie & Rayner, 1975; Rayner, 1975). However, there is debate about the extent and the quality of the parafoveal information that is available when the oculomotor system makes a decision about where to fixate next and when to make the next saccade (see Liversedge & Findlay, 2000; Rayner, 1998, 2009; Rayner & Liversedge, 2011; Schotter, Angele, & Rayner, 2012 for reviews)[1]. Second, readers can also use the preceding sentence context to predict/guess (at an unconscious level) the identity of the upcoming words (Balota, Pollatsek, & Rayner, 1985; Bicknell & Levy, 2010; Ehrlich & Rayner, 1981; Levy, Bicknell, Slattery, & Rayner, 2009; Rayner & Well, 1996). However, such predictions/guesses may not always be correct and may require input from high-level sentence integration processes which might not be available early enough to influence oculomotor decisions. The question we will consider is to what extent readers use parafoveal information and sentence context information when they make a skipping decision.

In determining the relative importance of parafoveal information compared to predictability from the sentence context, the phenomenon of word skipping is particularly interesting: if a word is intentionally skipped, this can be taken to imply that it has been parafoveally processed to the point where it no longer needs to be fixated. This assumption is explicitly implemented in serial attention shift models of eye-movement control during reading such as the E-*Z* Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Warren, & McConnell, 2009), according to which a word is never intentionally skipped unless it has been parafoveally identified. This view derives from the assumption that words are processed one at a time and in the correct order (but not necessarily fixated in that order). Word skipping in E-*Z* Reader occurs because readers start preprocessing the upcoming word parafoveally whenever they have identified the currently fixated word but their oculomotor system is not yet ready to perform a saccade to the next word. If the upcoming word is very easy to process, it is possible that it too is identified before the saccadic program has completed. In this case, the current saccadic program is cancelled and a saccade to the second word to the right—that is, a skipping saccade—is prepared[2].

A different approach to word skipping – also stressing the importance of parafoveal information – is made by processing gradient models such as the SWIFT model of eye-movement control (Engbert, Nuthmann, Richter, & Kliegl, 2005), which assume that multiple words can be processed in parallel and therefore allows for word skipping even if a word has not been fully identified. Specifically, SWIFT proposes a dynamic field of activations with one activation value for each word in a sentence. The maximum activation value of each word is determined by its word frequency. Initially, the further a word has been processed, the higher its activation value is (predictable words rise in activation more slowly than unpredictable words). Once the activation of a word reaches its maximum, further processing lowers the activation value until it is back at zero at which point the word is assumed to be fully identified. In SWIFT, saccades are triggered at random intervals (with a possible delay due to foveal word difficulty). The probability of a word becoming the target of the next saccade is directly proportional to its activation level relative to the activation levels of the other words in the field. This means that any word that has been processed to any degree and that is not yet fully identified can be a potential saccade target in SWIFT. Because of this, skipping the upcoming word is possible as soon as the word after it has accumulated some activation. For example, if the upcoming word has a relative activation of .7 and the word to its right has a relative activation of .2, SWIFT will skip the upcoming word in 20% of simulations (and fixate it in 70% of simulations).

Importantly, both E-Z Reader and SWIFT allow for an influence of word predictability from the sentence context on skipping probability. However, neither model is very clear on how the oculomotor system estimates predictability. In particular, both models assume that, in order for predictability to play a role in word identification, the word in question must have received at least some parafoveal preprocessing – simply guessing a word without any parafoveal input does not occur according to either model. Finally, both the SWIFT and E-Z Reader models can also account for accidental word skipping due to oculomotor error.

Much of the research on word skipping has focused on those words that are skipped most often during reading, such as the definite article *the* in English or the plural definite article *les* in French. O'Regan (1979) and Gautier, Le Gargasson, and O'Regan (2000) found that *les* was skipped more often than other three-letter words even when it was not predictable from the prior context. Angele and Rayner (2012) investigated this "automatic" *the*-skipping phenomenon further by using the gaze-contingent boundary paradigm (Rayner, 1975) to present readers with infelicitous (syntactically illegal) previews of the article *the*. This was done by having subjects read sentences containing three-letter target verbs (e.g. *ate*). In the critical condition, while fixating to the left of the target word, the parafoveal preview for the target verb was replaced with the article *the*. This ensured that all *the*-previews were infelicitous in the sentence context. There were two additional conditions, one in which the preview was correct (control condition) and a random letter nonword preview condition.

After crossing the boundary, readers always saw the correct word. Angele and Rayner found that readers skipped nearly 50% of the words that appeared as *the*, even when the sentence context did not syntactically allow an article in the target word position. This skipping rate was comparable with that calculated for felicitous occurrences of *the* in other positions in the sentences, suggesting that, in accordance with Gautier et al.'s results, skipping of articles is not strongly influenced by constraints of the preceding sentence context. In other words, Angele and Rayner's results seem to indicate that word skipping is mainly influenced by parafoveal processing. However, it is not clear whether article previews are skipped more often than correct verb previews because of their higher word frequency alone, or whether the lower fixation probability is specific to the article *the* (and perhaps a small number of other common, short function words).

In order to investigate this important theoretical question further, we used a paradigm that was quite similar to the one used by Angele and Rayner (2012). Specifically, we used three-letter target words (*dog* in the sentence "The excitable dog was ready to go for his walk"), the previews of which could either be correct and compatible with the sentence context (identical control condition, *dog*), a random letter string (*fze*), or a higher- or lower-frequency three-letter word that did not fit in the sentence context (*dim*). According to the CELEX corpus (accessed using the *N-Watch* software by Davis (2005), the low frequency targets had a mean word frequency of 10 (SD = 14), while the high frequency words had a mean word frequency of 1177 (SD = 3110). The mean difference in frequency within each pair was 1168 (SD = 3108; see Table A2 in the Appendix for details). It is important to note that, in the above example, presenting an adjective preview instead of the target word (which was a noun) caused a syntactic violation. This was the case in the majority of our stimuli (see the Method section for details). Due to their limited horizontal extent, such short words should be skipped quite often in reading (Hautala, Hyönä, & Aro, 2011; Rayner & McConkie, 1976). On the other hand, high frequency words are skipped more often than low frequency words (White, 2008). As a consequence, we expected there to be a number of potential influences on fixation probability and fixation times.

First, the sentence context is very likely to have a strong influence on all eye-movement measures. However, it is important to note that the main effect of sentence frame is not readily interpretable as the pre- and post-target words necessarily differed between sentence frames. Even on the target word, the effect of the context might be influenced by spillover effects from the previous words. Consequently, in addition to the main analysis across sentence frames, we will present separate analyses for those sentence frames in which the target was a high-frequency word (and the dissimilar previews could either be a random letter string or a lower frequency word) and those sentence frames in which the target was a low-frequency word (and the dissimilar previews could be a random letter string or a higher frequency word).Within sentence frames, the context was constant across all preview conditions.

Second, the parafoveal preview should have a clear effect on the processing of the target word. Strong preview benefit effects on the target word were predicted, that is, shorter fixation times when the target word preview was identical to the target word and longer fixation times when it was dissimilar. We also anticipated that these effects might spill over onto the post-target word.

Finally, we expected the parafoveal preview to have an immediate effect on eye movement behavior before fixating the target word: If the preview is a non-word, subjects should be more likely to fixate the target word. The presence of a non-word in the parafovea may also lead to longer fixation times on the pre-target word (such an effect of parafoveal orthographic information on ongoing processing as evidenced by the duration of the current fixation is known as an orthographic parafoveal-on-foveal effect; see Schotter et al., 2012). For those conditions in which the preview is a word (that is, the identical and alternative preview conditions), the probability of skipping the target word could be affected by the frequency of the preview and by the fit between preview and sentence context. If readers only take the frequency of the upcoming word into account, the frequency preview should have a direct influence on whether the target word is skipped or fixated independent of the sentence context, with high-frequency previews leading to higher skipping rates than low-frequency previews. Alternatively, if readers take the context into account when making a skipping decision, they should be less likely to skip a syntactically invalid preview than a syntactically valid preview. Of course, target word skipping might be affected by both preview word frequency and contextual fit at the same time. Additionally, based on previous studies (Kennedy, 1998; Kennedy, Murray, & Boissiere, 2004; Kennedy & Pynte, 2005; Kennedy, Pynte, & Ducrot, 2002; Kliegl, Nuthmann, & Engbert, 2006), a low-frequency word in the parafovea might also lead to longer fixation times on the pre-target word than a high-frequency word (a lexical parafoveal-on-foveal effect). In more general theoretical terms, the outcome of this experiment will tell us more about the importance of predictive linguistic processing during reading.

Method

Subjects

Thirty-five students from the University of Southampton participated in the experiment for course credit. All were native English speakers with normal/corrected-to-normal vision and no diagnosed reading difficulties.

Apparatus

Eye-movements were monitored every millisecond using an SR Research Eyelink 1000 eye-tracker. Viewing was binocular, though eye movement data were only collected from the right eye. Sentence stimuli were displayed on a computer monitor with a refresh rate of 150 Hz[3]. Viewing distance was 73 cm, with 3 characters equaling 1° of visual angle. A video-game controller was used by subjects to end each trial and respond to comprehension questions.

Materials

We assembled a list of 120 pairs of three-letter high- and low-frequency words; members of each pair differed in word frequency. Furthermore, we constructed 240 sentence frames (see Appendix for details). Each sentence frame was compatible with either the high or the low frequency word of a pair, but not with the other word of that pair. For most of the word pairs and sentence frames, this incompatibility was syntactic (144 out of 240), but, due to the difficulty inherent in finding word pair/sentence frame combinations resulting in true syntactic anomalies, about one third of the word pair/sentence frame combinations instead had semantic incompatibilities (*It was an emotional day/dew for all of the family members*, 96 out of 240). Some of the target word pairs also contained content words (e.g. *can/cow*). We will present a number of supplementary analyses aimed at determining whether these differences between items had an impact on the effect of the preview manipulation at the end of the next section. Each subject read only one of the two sentence frame versions, resulting in an equal number of high-to-low and low-to-high frequency manipulations (see Figure 1 for an example).

Insert Figure 1 about here

To summarize, there were three preview conditions: the preview word was either identical to the target word *(dog/dog; dim/dim*), a dissimilar random letter-string *(fze/dog; fzj/dim*), or the infelicitous lower/higher frequency alternative word *(dim/dog; dog/dim)*. Each subject read a total of 120 sentences.

Procedure

Subjects were asked to read each sentence for comprehension. A chin/forehead rest was used to minimize participant head movements. Initial three-point calibrations were carried out until error was <0.3° and re-calibrations were completed as needed. The entire session lasted approximately 45 minutes[4].

Results and Discussion

We computed a number of standard eye movement measures (Rayner, 1998, 2009) on the pre-target word (e.g. *excitable*), the target word (*dog*), and the post-target word (*was*). Specifically, we computed two early processing time measures, *first fixation duration* (FFD; mean duration of the first fixation on a word) and *gaze duration* (GD; the sum of all fixations on a word before leaving it), both calculated only for words that were not initially skipped. Most importantly, we computed the *probability of skipping the target word* as well as the *probability of making a regression out of the target word*. The latter measure only showed significant effects on the post-target word. We therefore do not report the results from the regression probability analyses in the target and the pre-target word regions. Finally, we also calculated two later processing measures, *go-past time* (Go-past; the sum of all the fixations on a word and any regressive fixations before moving to the right of the target word) and *total time* (TT; the sum of all fixations on a word during a trial). Tables 1a through 3a show the means and standard deviations for each condition and each of the dependent variables.

In the present study, 28.6 % of the display changes completed more than 10 ms after the onset of the subsequent fixation. In this case, we discarded the corresponding trial from the analysis. This was done because previous research (Slattery, Angele, & Rayner, 2011) indicated that display changes delayed by more than 10 ms cause a change in eye-movement behavior even when subjects are not aware of them. Furthermore, if a fixation was shorter than 80 ms and located within one character space (11 pixels) of another fixation, it was merged into that fixation, otherwise it was deleted. Fixations with durations that deviated from a subject's mean by more than two standard deviations were deleted as well (around 5% of the data). All subjects answered at least 85% of the comprehension questions correctly.

We expected that the frequent skipping of the three-letter target words and exclusion of delayed display changes would lead to unequal cell sizes, which would make the use of ANOVAs to analyze the data difficult. Thus, we used linear mixed models (LMM) with the target word preview condition (identical vs. alternative vs. dissimilar) and the sentence frame used (high frequency target word vs. low frequency target word) as well as their interaction as fixed effects; additionally, the model contained random intercepts and random slopes for preview condition and sentence frame condition (Baayen, Davidson, & Bates, 2008)[5]. For the preview condition, we used successive differences contrasts, comparing the identical with the alternative and the alternative with the dissimilar condition. Since the sentence frames differed in many ways, interpreting the main effect of sentence frame is not possible. However, the interaction of sentence frame with preview is still interesting, especially on the target word. In order to further investigate cases where this interaction reached significance, we performed separate analyses for each of the two sets of sentence frames. We used the lmer function from the lme4 package (Bates, Maechler, Bolker, & Walker, 2013) within the R Environment for Statistical Computing

(R Development Core Team, 2013) to fit the LMMs. For each factor (preview word, sentence frame, and their interaction), we report regression coefficients (*b*), standard errors, and *t*-values. For binomial dependent variables such as fixation and regression probabilities, we report regression coefficients, standard errors, and *z*-values from generalized LMMs using a logit-link. It is not clear how to determine the degrees of freedom for the *t*-statistics estimated by the LMMs, making it difficult to estimate *p*-values (Baayen et al., 2008). However, since our analyses contain a large number of subjects and items and only a few fixed and random effects are estimated, we can assume that the distribution of the *t*-values estimated by the LMMs approximates the normal distribution. We therefore used the two-tailed criterion $|t| \geq 1.96$ which corresponds to a significance test at the 5% α-level. The *z*-values from the generalized LMMs can be interpreted in exactly the same way. Tables 1b through 3b (sentence frames with high-frequency target word) and 1c through 3c (sentence frames with low-frequency target word) show the LMM results, although the coefficient estimates and statistics for significant effects are also repeated in the text.

Insert Tables 1-3 around here

Early processing measures: Skipping probability

Pre-target word. The probability of skipping the pre-target word was influenced by the preview condition; readers were significantly less likely to skip the pre-target word in the alternative condition than in the identical condition ($b = .45$, SE = .18, $z = 2.49$). Readers were also less likely to skip the pre-target word in the dissimilar preview condition than in the alternative preview condition ($b = -.44$, SE = .19, $z = -2.3$). These effects, which are similar to effects observed by Angele and Rayner (2012), were unexpected and it is not quite clear what causes them. Perhaps readers are on some occasions able to detect an anomaly arising from the preview manipulation on the target word while fixating two words to the left of it (see Rayner, Angele, Schotter, & Bicknell (2013) for discussion), though a fair amount of prior research indicates this is unlikely (see Rayner, 1998, 2009 for reviews). Consistent with our last point, note, however, that skipping of this region only occurred on a minority of trials (between 11 and 17%)—in the majority of trials, readers either do not notice the parafoveal violation or do not react to it until they reach the pre-target word.

Target word. Both the difference in skipping probability between the identical and alternative conditions (b = .52, SE = .21, z = 2.5) and the difference between the alternative and dissimilar conditions (b = -.61, SE = .21, z = -2.91) were modulated by sentence frame. Specifically, sentence frames with high-frequency targets showed a small, and, likely due to lack of power, only marginally significant difference between the alternative and the identical condition, with the target word being less likely to be skipped in the alternative than in the identical condition (b = -.29, SE = .17, z = -1.69), and barely any difference at all between the dissimilar and alternative conditions (b = .00, SE = .18, z = -.02). In contrast, sentence frames with low-frequency targets showed a small difference in the opposite direction between the alternative and the identical conditions, with the target being skipped less often in the identical than in the dissimilar condition (b = .25, SE = .15, z = -1.64), and a very strong difference between the dissimilar and the alternative conditions, with the dissimilar condition resulting in many more target fixations than the alternative condition (b = -.62, SE = .15, z = -4.17). This interaction suggests that, overall, readers were more likely to skip a target word with a high-frequency preview (the identical preview for high-frequency target sentence frames and the alternative preview for low-frequency target sentence frames) than a target word with a low-frequency preview (the alternative preview for high-frequency target sentence frames and the identical preview for low-frequency target sentence frames), confirming our prediction.

Post-target word. There was a significant interaction between sentence frame and preview on skipping probability on the post-target word. The difference in skipping probability between the alternative and the dissimilar preview conditions was modulated by sentence frame (b = -.46, SE = .23, t = -2.02). The separate analyses by sentence frame showed that, for high-frequency target sentence frames, there was a marginally significant difference between the alternative and the dissimilar preview conditions, with the dissimilar preview leading to more skips of the post-target word (b = -.33, SE = .18, z = -1.86). For low-frequency target sentence frames, on the other hand, this effect was completely absent (z = .9)[6]. In summary, skipping of the post-target word was not strongly affected by the preview manipulation.

Early processing measures: Fixation time measures

Pre-target word. There were no significant effects of preview, with the exception of a significant interaction between the preview contrast between the identical and the alternative preview and sentence frame in FFD (b= -14.18, SE = 5.73, t = -2.48). However, it is unclear whether this effect can be interpreted[7]. There were no significant effects on other early fixation time measures on the pre-target word (all ts< 1.96).

Target word. In the early fixation time measures we found evidence of a standard preview benefit effect (Rayner, 1975): There was a main effect of preview in FFD and GD indicating that fixation times on the target word were shortest in the identical condition compared to the alternative condition (FFD: b = 15.53, SE = 4.94, t = 3.14; GD: b = 19.91, SE = 7.61, t = 2.62). Furthermore, FFD and GD were shorter in the alternative condition than in the dissimilar condition (FFD: *b* = 13.23, SE = 4.93, *t* = 2.68, GD: *b* = 17.34, SE = 6.86, *t* = 2.53), suggesting that there was a cost of having a dissimilar, random letter preview that exceeded the cost of having a preview that was simply a different word. On GD, this effect differed between sentence frames (Interaction term: *b* = -25.47, SE = 12.59, *t* = -2.02). Specifically, there was a strong difference between the alternative and the dissimilar condition in sentence frames with a high-frequency target (*b* = 29.12, SE = 8.43, *t* = 3.46) and virtually no difference in sentence frames with a low-frequency target (b = 4.44, SE = 10.95, *t* = .4). Apparently, processing a low-frequency target word was much more dependent on a correct preview than processing a high-frequency target word. The dissimilar preview proved disruptive in both cases.

Post-target word. There was a spill-over effect of the target word manipulation. Specifically, GD on the post-target were significantly longer in the alternative preview condition compared to the identical condition (GD: *b* = 22.69, SE = 7.68, *t* = 2.95). This effect did not differ between sentence frames (*t* < 1.96).

Late processing measures

Pre-target word. There was a main effect of preview on total viewing time in those sentence frames with a high-frequency target word, with a significant difference in TT between the identical high-frequency target previews and the alternative low-frequency word previews (*b* = 26.35, SE = 11.44, *t* = 2.3). There also was a significant interaction between preview (alternative vs. identical contrast) and sentence frame (*b* = -35.65, SE = 16.89, *t* = -2.11). Separate analyses by sentence frame (see Table 1c) showed a significant difference between the alternative and the identical conditions in sentence frames with a high-frequency target word (*b* = 46.99, SE = 15.11, *t* = 3.11), but not in those with a low-frequency target word (*b* = 7.66, SE = 16.14, *t* = .21). Overall, preview effects for TT on the pre-target word are most likely due to readers re-reading the pre-target word after having received a dissimilar preview. The fact that such an effect only reached significance for the sentence frames with a high-frequency target word may again indicate that processing was disrupted more when readers were expecting a high-frequency word but receive a dissimilar preview compared to the situation in which readers were expecting a low-frequency word and then receive a high-frequency or random letter preview. Of course, the preceding sentence context may have influenced the effect of a dissimilar preview in other ways as well.

Target word. The preview benefit effect we found on the earlier measures persisted in the late measures, with longer go-past times and TT in the alternative than in the identical condition (go-past: $b = 46.29$, SE = 13.2, $t = 3.51$; TT: $b = 28.23$, SE = 8.71, $t = 3.24$).

Post-target word. We found spillover effects on go-past time and TT (go-past: $b = 64.87$, SE = 14.9, $t = 4.35$; TT: $b = 28.67$, SE = 9.81, $t = 2.92$), with go-past times and TTs being longer in the alternative preview condition than in the identical condition. Again, this effect was not modulated by sentence frame ($b = -2.75$, SE = 18.41, $z = -.15$).

Supplementary analyses

Syntactic vs. semantic preview violations. As described in the Method section, about one third of the sentence frame/preview word combinations resulted in a semantic rather than syntactic violation (e.g. *They looked at the afternoon (sky/cut) and admired the clouds* is a semantic preview violation as opposed to the syntactically illegal preview violation in *She accidentally (cut/sky) herself while making a collage*, where the first word in parentheses is the target word and the second word in parentheses is the alternative preview). In order to determine whether the observed effects differed by type of violation, we performed analyses with an additional factor denoting whether the violation in the alternative preview condition was syntactic (coded as 0) or semantic (coded as 1, see Appendix for details). There was no significant interaction effect on skipping probability between violation type and preview and no significant three-way interaction between preview, sentence frame, and violation type (all $z < 1.65$), suggesting that the effect we observed did not differ between violation types. There were a small number of significant interaction effects on fixation time measures. Specifically, there was a significant interaction between violation type and the comparison between the identical and the alternative previews in FFD and GD on the post-target word (FFD: b= -16.67, SE = 8.24, t = -2.02; GD: b = -31.87, SE = 14.7, t = -2.17) when the alternative preview constituted a syntactic violation, suggesting that the spillover effect from having had an alternative as opposed to an identical preview of the target word was only present when the violation had been of a syntactic nature. There was also a significant interaction between violation type and the comparison between the alternative and the dissimilar previews in GD and go-past time on the target word (GD: b = -28.42, SE = 12.97, t = -2.192; Go-past: b = -64.15, SE = 24.67, t = -2.60), indicating that, in contrast to the main analysis, there was a difference in preview benefit between the alternative and the dissimilar conditions, but only when the preview violation was of a syntactic nature. However, as the present study was not designed to investigate the effect of preview violation types and our analyses may not have enough power to detect subtle effects of violation type, further research will be necessary in order to confirm that, for the purposes of making a skipping decision, semantic and syntactic preview violations are equivalent (that is, that neither type has an effect on skipping).

Content vs. function word targets and previews. Another factor that may be important in the present study is the use of both content and function words as targets and alternative previews (e.g. *The old man across the street (has/hem) a very bad cold* vs. *My grandmother told me she would (hem/has) that dress for me*, where the first word in parentheses is the target word and the second word in parentheses is the alternative preview). The advantage of using function words as targets and previews is that this is likely to result in a strong syntactic violation. However, since one of the goals of the present study was to determine whether skipping a preview consisting of a syntactic violation occurs only for function word previews, it is important to check whether the effects we observed were driven entirely by such target and preview words. In order to do this, we performed analyses with an additional factor denoting whether both target and preview words were content words (coded as 0) or whether either the preview or target word was a function word (coded as 1, see Appendix for details); there were no cases in which both preview and target were function words. There was no significant interaction effect between preview and target/preview word class on any of the skipping probabilities for the pre-target, target, and post-target words and no significant three-way interaction between preview, sentence frame, and target/preview word class (all $z < 1.96$), with one exception: the post-target word was skipped more often when the preview had been dissimilar (as opposed to alternate) when the target word was a function word ($b = .71$, SE = .25, $z = 2.8$). Importantly, the effect we observed on the probability of skipping the target word did not differ depending on whether target and preview words were content or function words. There was only one significant interaction effect on fixation time measures. Specifically, there was a significant three-way interaction on go-past time on the target word between the identical vs. alternative preview contrast, the sentence frame, and target word class ($b = -130.11$, SE = 60.08, $t = -2.17$), suggesting that the preview benefit effect on go-past time on the target word was only present for those sentence frames in which the target word was a high-frequency function word. None of these interactions provide evidence that detracts from our primary claim that high frequency parafoveal words are skipped more often than low frequency parafoveal words even if the sentence context favors the low frequency word. Furthermore, based on these analyses, it appears that the content or function status of a word did not modulate these effects.

Target word predictability. Finally, the target words differed in terms of their predictability from the preceding sentence context. While the majority of the target words were not predictable from the context, a number of target words were somewhat constrained by the sentence context. We therefore collected target word cloze data for all sentence frames from 11 UK native speakers in order to assess predictability. Predictability was low for most, but not all of the target words with a few exceptions (mean = .07; range = 0-.91) In order to assess whether predictability modulated the effects we observed, we performed another set of analyses in which we included target word predictability in the analysis. Given that there were 11 subjects, we entered target word predictability into the model as a categorical rather than a continuous variable (target word or target word preview were predicted by at least one norming subject, coded as 1 –this included 29 % of items with an average predictability of .23–vs. neither target word nor preview were predicted by any norming subjects, coded as 0—this included 71% of items; see Appendix for details). There were no significant two- or three-way interaction effects of target word predictability on the probability of skipping the pre-target, target, or post-target word, suggesting that target word predictability did not affect subjects' skipping decisions in the present study (all $z < 1.96$). There were a few significant interaction effects on fixation time measures: On the pre-target word, there were significant three-way interactions between target word predictability, sentence frame, and each of the preview contrasts (interaction term for identical vs. alternative: b = -34.08, SE = 13.63, $t = -2.5$; alternative vs. dissimilar: b = 29.94, SE

= 13.36, $t = 2.24$). This is quite interesting, as it suggests that there was a parafoveal-on-foveal effect of preview when the target word was predictable from the sentence context, but no parafoveal-on-foveal effect of preview when it was not predictable. Separate analyses for the sentences with predictable target words and those with non-predictable target words produced interactions between the preview contrasts and the sentence frame that were not significant for those sentences in which the target word was not predictable (all *t*s < 1), but highly significant for those sentences in which it was. However, this effect was only significant for the sentence frames compatible with high-frequency target words (Interaction term for identical vs. alternative: b = -42.02, SE = 11.43, $t = -3.68$; alternative vs. dissimilar: b = 30.72, SE = 11.21, $t = 2.74$). Therefore, it appears that subjects may have, at least in some cases, predicted the identity of the target word, and when the target preview was incompatible with the prediction, they experienced some disruption. A similar effect was observed by Rayner et al. (2013). Such a prediction mechanism could also explain the apparent parafoveal-on-foveal effects we observed for skipping the pre-target word, although this would suggest that both prediction generation and comparison of the prediction with the parafoveal preview can occur as much as two words ahead. Such a suggestion might occur if the perceptual span is extended when upcoming words are highly predictable. Finally, the interaction between the identical vs. alternative preview contrast and word predictability on FFD on the post-target word was significant (b = -20.17, SE = 9.54, $t = 2.12$), suggesting that, for predictable target words, there was no spillover effect of having had the alternative preview—on the contrary, FFDs on the post-target word were slightly lower in the alternative preview condition compared to the identical condition.

General Discussion

In the present experiment, we pitted two sources of information about the upcoming word against each other. In the critical preview condition, the information from the available parafoveal preview contradicted expectations about the upcoming word based on the syntactic structure of the preceding sentence. Angele and Rayner (2012) showed that, in cases when the preview indicates that the upcoming word is the article *the*, but the sentence context is incompatible with that conclusion, the article interpretation wins: readers skip the apparent article and incur substantial processing difficulties later on. The present experiment demonstrates that a similar effect occurs for other three-letter words: First, random letter non-words were skipped less often than words. Second, the frequency of the upcoming word, but not its fit with the sentence, determined whether it was skipped. This finding is also consistent with the findings of White (2008), who showed that high frequency words are usually skipped more often than low-frequency words.

The present findings suggest that, at least for very short words, lexical properties of the parafoveal preview have a substantial influence on fixation probability, while its compatibility with the preceding sentence context matters to a far lesser degree. Additionally, properties of the preview affect later processing, as shown by higher regression probabilities and go-past times. Our findings also suggest that readers are able to obtain some lexical information about parafoveal words, which fits neatly with the fact that readers seem to be able to skip even difficult words and still comprehend a sentence without problems. On the other hand, while expectations based on the preceding sentence context certainly play a role in shaping reading behavior (see Bicknell & Levy, 2011), parafoveal lexical processing seems to trump the sentence context when it comes to word skipping.

A remaining question is how our results relate to previous findings about the influence of predictability on word skipping (Balota et al., 1985; Drieghe, Rayner, & Pollatsek, 2005; Ehrlich & Rayner, 1983; Rayner & Well, 1996; for a review see Brysbaert, Drieghe, & Vitu, 2005), which suggest that readers are more likely to skip highly predictable words. Given this, why did we not see more pronounced effects of predictability in the present study? As noted above, this likely occurred because our target words were, in general, not strongly predictable (see Appendix for details). It also remains to be determined whether our results generalize to longer words (for a recent study on the skipping of longer words, see Choi & Gordon, 2013).

In summary, the present study demonstrates that skipping of short words is strongly influenced by the frequency of its parafoveal preview. This suggests that the *the*-skipping effects observed by O'Regan, (1979), Gautier et al. (2000), and Angele and Rayner (2012) are not necessarily specific to the definite article *the*, but rather, they occur as a function of its word frequency. On a more general level, our results underline the importance of parafoveal orthographic and lexical processing in comparison with higher-level processing such as syntactic or semantic processing. While syntactic integration and semantic processing occur during or after a word has been fixated, the information available about a word that is still in the parafovea is mainly of either an orthographic or lexical nature. Importantly, this does not mean that parafoveal information about a word is irrelevant once the eyes have moved—on the contrary, we found clear effects of the preview of a word having been dissimilar once that word is finally fixated and even on the subsequent fixations. Clearly, however, there is no evidence with this experimental paradigm that syntactic and semantic information about the upcoming word and its fit with the sentence affect the timing of the skipping decision.

Implications for computational models

Our findings show that readers routinely skip parafoveal high frequency words without taking the syntactic or semantic sentence context into account. This demonstrates that, at least with regards to skipping short words, the current implementations of both the E-Z Reader and SWIFT models are adequate: In both models, the time needed to parafoveally process a word (and its fixation probability) is strongly dependent on its frequency. Both models also predict that the processing time of a word should be influenced by its predictability, however, this is where our results do not correspond to the model predictions, as we saw no effect of fit with the sentence context. It is important to note, however, that both E-Z Reader 9 (Pollatsek, Reichle, & Rayner, 2006), and SWIFT focus on word identification rather than semantic integration. As such, they both allow parafoveal identification of an upcoming word prior to skipping it. Version 10 of E-Z Reader (Reichle et al., 2009) goes further than this by including a processing stage representing the syntactic and semantic integration of a word into the sentence context. If one assumes that the integration stage in E-Z Reader 10 would necessarily fail to integrate the infelicitous *the* preview into the sentence, E-Z Reader 10 might be able to explain why there was disruption only subsequent to skipping the infelicitous *the* preview. Keeping in mind that there are previous results supporting an effect of predictability on word skipping and that our experiment really only measures the difference in skipping between an unpredictable and grammatically illegal (or, in the case of the semantic violations, at least implausible, if not anomalous) parafoveal word, some changes in how the models treat the predictability of parafoveal words might be necessary. However, such changes would be rather small modifications of existing mechanisms in both models and would not require that new mechanisms be proposed. Furthermore, these effects can be accounted for by both parallel and serial accounts of word identification.

The unexpected parafoveal-on-foveal effect of the preview condition on skipping probability and fixation time on the pre-target word might have stronger implications for the models. If there is indeed an early parafoveal plausibility check, such a mechanism is not present in either model and would have to be added. One might argue that processing far ahead of the fixated word may be more in the “spirit” of parallel processing, however, it would not be impossible to conceive of a serial model that involves a very early parallel visual stage in which all letters in the perceptual span are superficially processed at the same time (in fact, the “V” stage in E-Z Reader is described as exactly such a process, though this is not implemented in current versions of the model). This superficial processing could involve checking whether parafoveal word information is consistent with those words that are pre-activated from processing the context. It is important to note at this point that, in order to warrant model modifications, the parafoveal-on-foveal effects described above will have to be replicated in a dedicated experiment. They do suggest some interesting directions for future research, however.

In summary, our finding that high frequency parafoveal words are often skipped without taking the sentence context into account is, in principle, compatible with both E-Z Reader and SWIFT (pending minor modifications). Explaining our findings on the pre-target word might require more fundamental model changes, but further research is required to replicate those effects in a controlled experiment. On a more general level, our results indicate that there is at least one important process during reading which does not seem to be affected by word predictability from the context, but just by the properties of the parafoveal word. This may pose a problem to models, such as that of Bicknell and Levy (2010), that propose that readers constantly evaluate the available evidence about previously fixated, currently fixated, and upcoming words since such an evaluation should alert readers to the mismatch between prediction and actual parafoveal information. While we did find some evidence for such a parafoveal plausibility check, the effects we observed were not strong enough to warrant the assumption that such a check is carried out during every fixation. Rather, it seems that, during normal reading, parafoveal plausibility is only checked once in a while if at all.

References

Angele, B., & Rayner, K. (2013). Processing *the* in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(2)*, 649-662. doi:10.1037/a0029294

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, *17*, 364–390.

Bates, D., Maechler, M., Bolker, B. & Walker S. (2013). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5.

http://CRAN.R-project.org/package=lme4

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 1168–1178).

Bicknell, K., & Levy, R. (2011). Why readers regress to previous words: A statistical analysis. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society. (pp. 931–936). Austin, TX: Cognitive Science Society.

Brysbaert, M., Drieghe, D., & Vitu, F. (2005). Word skipping: Implications for theories of eye movement control in reading. In G. Underwood (Ed.), Cognitive processes in eye guidance (pp. 53-77). Oxford: Oxford University Press

Choi, W., & Gordon, P. C. (2013). Coordination of word recognition and oculomotor control during reading: the role of implicit lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(4), 1032–1046. doi:10.1037/a0030432

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 954-969.

Davis, C. J. (2005). N-watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*(1), 65-70.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, *20*, 641–655. doi:10.1016/S0022-5371(81)90220-6

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*, 777–813. doi:10.1037/0033-295X.112.4.777

Gautier, V., O’Regan, J. K., & Le Gargasson, J. F. (2000). “The-skipping” revisited in French: programming saccades to skip the article “les”. *Vision Research*, *40*, 2517–2531.

Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research*, *51*, 1719–1727. doi:10.1016/j.visres.2011.05.015

Kennedy, A. (1998). The influence of parafoveal words on foveal inspection time: Evidence for a processing trade-off. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 149–179). Oxford, UK: Elsevier.

Kennedy, A., Murray, W., & Boissiere, C. (2004). Parafoveal pragmatics revisited. *European Journal of Cognitive Psychology*, *16*, 128–153.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*, 153–168.

Kennedy, A., Pynte, J., & Ducrot, S. (2002). Parafoveal-on-foveal interactions in word recognition. *The Quarterly Journal of Experimental Psychology Section A*, *55*, 1307–1337.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology-General*, *135*, 12–35.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*, 21086–21090. doi:10.1073/pnas.0907664106

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, *4*, 6–14. doi:10.1016/S1364-6613(99)01418-7

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, *17*, 578–587.

O'Regan, K. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, *25*, 501–509. doi:10.3758/BF03213829

Pollatsek, A., Reichle, E.D., & Rayner, K. (2006). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*, 1-56.

R Development Core Team, R. (2013). R: a language and environment for statistical computing. *R Foundation for Statistical Computing*.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, *7*, 65–81. doi:10.1016/0010-0285(75)90005-5

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K. (2009). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506. doi:10.1080/17470210902816461

Rayner, K., Angele, B., Schotter, E. R. and Bicknell, K. (2013). On the processing of canonical word order during eye fixations in reading: Do readers process transposed word previews? *Visual Cognition*, 21 (3), 353-381.

Rayner, K., & Liversedge, S. (2011). Linguistic and cognitive influences on eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements*. Oxford, UK: Oxford University Press.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research*, *16*, 829–837. doi:10.1016/0042-6989(76)90143-7

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, *3*, 504–509.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, *105*, 125–157. doi:10.1037/0033-295X.105.1.125

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, *16*, 1–21. doi:10.3758/PBR.16.1.1

Schotter, E.R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception, & Psychophysics*, *74*, 5–35. doi:10.3758/s13414-011-0219-2

Slattery, T. J., Angele, B., & Rayner, K. (2011). Eye movements and display change detection during reading. *Journal of Experimental Psychology. Human Perception and Performance*, *37*, 1924–1938. doi:10.1037/a0024322

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception & Psychophysics*, *57*(3), 352–364. doi:10.3758/BF03213060

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 205–223. doi:10.1037/0096-1523.34.1.205

White, S., Rayner, K., & Liversedge, S. (2005). Eye movements and the modulation of parafoveal processing by foveal processing difficulty: A reexamination. *Psychonomic Bulletin & Review*, *12*, 891–896. doi:10.3758/BF03196782

Acknowledgments

This research was supported by Leverhulme Trust Grant F/00 180/AN, Economic and Social Research Council Grant ES/I032398/1 and by National Institutes of Health Grant HD26765. We thank Emily Morgan and Roger Levy for their help with collecting the cloze norming data, and Reinhold Kliegl for comments on a prior draft.

Footnotes

[1] An extreme position on this question would hold that skipping is almost entirely explainable by oculomotor factors, including word length (for an example, see Vitu, O'Reagan, Inhoff, & Topolski, 1995). However, this fails to explain how word frequency and predictability can have an effect on word skipping as detailed below.

[2] In rare cases, the oculomotor error built into E-Z Reader can lead to accidental skipping of a word. However, in this case, skipping the word would not be considered intentional.

[3] At a refresh rate of 150 Hz, the display changes took an average of 3 ms and a maximum of 6 ms to be completed after they were initiated. As a consequence, after removing those trials with very late display changes (see Results and Discussion section), the average display change finished less than 2ms after fixation onset.

[4] There is evidence that awareness of display changes can influence some eye-movement measures (White, Rayner, & Liversedge, 2005). Out of 53 subjects that were originally tested, the data of 18 subjects who noted display changes on their own without prompting or noticed more than 5 changes were excluded from the analysis because of this (leaving a total of 35 subjects included in the analysis). An analysis of the data from the 18 discarded subjects showed few differences in terms of the effects pattern compared to the subjects in the main analysis; however, a number of effects failed to reach significance, likely due to power limitations. The skipping probability effects that did not reach significance for the discarded subjects were the

effect of preview (both contrasts) on the probability of fixating the pre-target word, the difference between the dissimilar and the alternative conditions in terms of the probability of skipping the target word as well as its interaction with sentence frame, and the effect of the interaction between the difference between the dissimilar and the alternative preview and the sentence frame on the probability of skipping the post-target word. The fixation time effects that did not reach significance for the discarded subjects were the interaction between preview and sentence frame on FFD on the pre-target word, the difference between the dissimilar and the alternative preview on FFD and GD on the target word, the interaction between preview and sentence frame on GD on the target word, the effect of sentence frame on FFD on the post-target word, on GD on the target word, and on go-past time on the post-target word, and, finally, the difference between the identical and alternative preview conditions on GD and go-past time on the post-target word. There was only one effect that was significant for the discarded subjects but not for the included subjects, namely an interaction between preview and sentence frame on GD on the post-target word (b = -76.82, SE = 26.19, t = -2.93).

[5] Random slopes for the interaction term could not be included, as the majority of the models no longer converged in this case. A small minority of models did not converge even with the restricted random effects specification described above. In this case, we report a more restricted model including random slopes only for the preview effect.

[6] There was an overall effect of preview on the probability of making a regression out of the post-target word, with more regressions in the alternative preview condition than in the

identical preview condition (b = .58, SE = .19, z = 3.11). This effect was not modulated by sentence frame and suggests that having had a parafoveal preview of a target word (or a dissimilar, random letter nonword) that does not fit in the later sentence context does disrupt further processing to some degree. Even though readers have seen the correct target word at this point, the preview information still seems to have some effect on their processing, possibly at the integration level.

[7] We are very cautious in interpreting this interaction effect, which is limited to FFD, for three reasons: (1) in order to detect the syntactic violation caused by the mismatch of sentence frame and preview, readers would have to have fully identified the preview and integrated it into the sentence frame before actually fixating the target word and (2) even if they did this, it is not clear why fixations would then be shorter in the syntactic violation condition rather than longer. Furthermore, (3) it is unclear why the syntactic violation would not also be detected in the high-frequency sentence frames as well.

Table 1a: Condition means on the pre-target word.

		Fixation time measures				Probabilities	
	Preview	**First fixation duration**	**Gaze duration**	**Go-past time**	**Total viewing time**	**Skipping probability**	**Probability of regressions out**
high	Identical	221 (57)	248 (94.7)	272 (133)	292 (153)	0.17 (0.38)	0.05 (0.22)
	Alternative	223 (57)	254 (104)	286 (173)	334 (193)	0.15(0.36)	0.06(0.24)
	Dissimilar	224 (58)	259 (112)	297 (192)	337 (207)	0.11 (0.31)	0.09(0.28)
low	Identical	227 (65)	261 (106)	307 (191)	319 (176)	0.12(0.32)	0.09 (0.2.9)
	Alternative	218 (56)	253 (114)	319 (254)	334 (217)	0.14(0.34)	0.1(0.3)
	Dissimilar	224 (59)	259 (125)	305 (218)	329 (183)	0.14(0.35)	0.08 (0.27)

Standard deviations are in parentheses.

Table 1b: LMM analyses on the pre-target word. Each column represents a model.

		First fixation duration			Gaze duration			Go-past time			Total viewing time			Skipping probability		
		B	**SE**	***t***	***b***	**SE**	***t***	***b***	**SE**	***t***	***b***	**SE**	***t***	***b***	**SE**	***z***
	(Intercept)	220.33	4.02	**54.74**	250.40	7.40	**33.83**	291.49	12.96	**22.50**	318.70	14.79	**21.55**	-2.57	0.22	**-11.84**
Preview	Alternative vs. Identical	-2.85	3.40	-0.84	-1.05	5.30	-0.20	9.50	10.75	0.88	26.35	11.44	**2.30**	0.45	0.18	**2.49**
	Dissimilar vs. Alternative	2.70	3.07	0.88	4.89	5.52	0.89	0.82	12.09	0.07	2.91	11.46	0.25	-0.44	0.19	**-2.30**
Sentence frame	High vs. Low	1.91	3.21	0.60	6.27	6.02	1.04	23.98	10.70	**2.24**	9.29	9.52	0.98	-0.30	0.17	-1.72
Interactions	Alternative vs. Identical × Sentence frame	-14.18	5.73	**-2.48**	-17.21	10.46	-1.65	-10.10	18.81	-0.54	-35.65	16.89	**-2.11**	0.58	0.32	1.81
	Dissimilar vs. Alternative × Sentence frame	5.72	5.64	1.01	1.01	10.36	0.10	-23.88	18.67	-1.28	-0.75	16.76	-0.04	0.47	0.31	1.49

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with |*t*| or |*z*| >= 1.96 are marked in bold.

Table 1c: LMM analyses on the pre-target word by sentence frame. Each column represents a model.

Sentence frames with high-frequency target words only		First fixation duration			Total viewing time		
		B	**SE**	***t***	***b***	**SE**	***t***
	(Intercept)	220.30	4.25	**51.84**	316.33	15.16	**20.87**
Preview	Alternative vs. Identical	3.48	4.84	0.72	46.99	15.11	**3.11**
	Dissimilar vs. Alternative	-0.28	4.08	-0.07	3.27	16.83	0.19

Sentence frames with low-frequency target words only		First fixation duration			Total viewing time		
		b	**SE**	***t***	***b***	**SE**	***t***
	(Intercept)	221.12	4.30	**51.39**	322.38	15.95	**20.22**
Preview	Alternative vs. Identical	-9.89	4.27	**-2.32**	7.66	16.14	0.47
	Dissimilar vs. Alternative	5.33	4.17	1.28	3.12	14.57	0.21

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with |*t*| or |*z*| >= 1.96 are marked in bold.

Table 2a: Condition means on the target word.

		Fixation time measures				Probabilities	
	Preview	**First fixation duration**	**Gaze duration**	**Go-past time**	**Total viewing time**	**Skipping probability**	**Probability of regressions out**
	Identical	235 (73)	250 (95)	283 (162)	275 (143)	0.52 (0.5)	0.05 (0.22)
high	Alternative	244 (67)	261 (95)	327 (192)	303 (155)	0.47 (0.5)	0.08 (0.28)
	Dissimilar	272 (80)	294 (99)	354 (199)	321 (141)	0.46 (0.5)	0.09(0.28)
	Identical	241 (69)	270 (107)	306 (165)	305 (156)	0.49(0.5)	0.048 (0.21)
low	Alternative	259 (80)	296 (112)	355 (211)	334 (166)	0.54(0.5)	0.07 (0.26)
	Dissimilar	266 (77)	300 (115)	354 (194)	332 (161)	0.41(0.49)	0.08 (0.28)

Standard deviations are in parentheses.

Table 2b: LMM analyses on the target word. Each column represents a model.

		First fixation duration			Gaze duration			Go-past time			Total viewing time			Skipping probability		
		B	**SE**	***t***	***b***	**SE**	***t***	***b***	**SE**	***t***	***B***	**SE**	***t***	***b***	**SE**	***z***
	(Intercept)	251.49	6.61	**38.03**	277.25	8.42	**32.91**	328.40	10.48	**31.33**	305.87	11.27	**27.13**	-0.03	0.15	-0.19
Preview	Alternative vs. Identical	15.53	4.94	**3.14**	19.91	7.61	**2.62**	46.29	13.20	**3.51**	28.23	8.71	**3.24**	-0.01	0.10	-0.08
	Dissimilar vs. Alternative	13.23	4.93	**2.68**	17.34	6.86	**2.53**	12.51	14.71	0.85	7.77	8.43	0.92	-0.31	0.11	**-2.72**
Sentence frame	High vs. Low	8.15	4.44	1.84	24.29	6.44	**3.77**	16.54	11.28	1.47	28.96	6.97	**4.16**	0.00	0.10	0.04
Interactions	Alternative vs. Identical × Sentence frame	6.35	9.25	0.69	11.79	13.00	0.91	-2.36	24.90	-0.09	-5.52	17.38	-0.32	0.52	0.21	**2.50**
	Dissimilar vs. Alternative × Sentence frame	-17.20	9.01	-1.91	-25.47	12.59	**-2.02**	-21.08	24.19	-0.87	-13.89	16.53	-0.84	-0.61	0.21	**-2.91**

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with |*t*| or |*z*| >= 1.96 are marked in bold.

Table 2c: LMM analyses on the target word by sentence frame. Each column represents a model.

Sentence frames with high-frequency target words only		Gaze duration			Skipping probability		
		b	**SE**	***t***	***b***	**SE**	***z***
	(Intercept)	266.51	8.19	**32.56**	-0.03	0.17	-0.18
Preview	Alternative vs. Identical	12.77	8.60	1.49	-0.29	0.17	-1.69
	Dissimilar vs. Alternative	29.12	8.43	**3.46**	0.00	0.18	-0.02

Sentence frames with low-frequency target words only		Gaze duration			Skipping probability		
		b	**SE**	***t***	***b***	**SE**	***z***
	(Intercept)	289.40	9.96	**29.06**	-0.07	0.14	-0.52
Preview	Alternative vs. Identical	25.52	11.73	**2.18**	0.25	0.15	1.64
	Dissimilar vs. Alternative	4.44	10.95	0.41	-0.62	0.15	**-4.17**

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with $|t|$ or $|z|$ >= 1.96 are marked in bold.

Table 3a: Condition means on the post-target word.

		Fixation time measures				Probabilities	
	Preview	First fixation duration	Gaze duration	Go-past time	Total viewing time	Skipping probability	Probability of regressions out
high	Identical	228 (69)	255 (106)	314 (199)	297 (154)	0.351 (0.48)	0.09(0.28)
	Alternative	235 (71)	275 (130)	381 (268)	326 (187)	0.324 (0.47)	0.15 (0.36)
	Dissimilar	226 (70)	271 (128)	386 (264)	335 (179)	0.279 (0.45)	0.17(0.38)
low	Identical	242 (76)	283 (139)	357 (217)	331 (180)	0.336 (0.47)	0.11 (0.31)
	Alternative	240 (76)	305 (140)	424 (259)	355 (180)	0.306 (0.46)	0.163 (0.37)
	Dissimilar	238 (75)	296 (137)	405 (254)	360 (187)	0.34 (0.47)	0.156 (0.364)

Standard deviations are in parentheses.

Table 3b: LMM analyses on the post-target word. Each column represents a model.

		First fixation duration*			Gaze duration			Go-past time			Total viewing time			Skipping probability			Probability of regressions out*		
		B	**SE**	***t***	***b***	**SE**	***t***	***b***	**SE**	***t***	***B***	**SE**	***t***	***b***	**SE**	***z***	***b***	**SE**	***z***
	(Intercept)	233.84	5.93	**39.46**	276.49	9.39	**29.45**	374.59	14.06	**26.64**	328.42	11.95	**27.49**	-0.92	0.11	**-8.78**	-2.09	0.14	**-14.55**
Preview	Alternative vs. Identical	4.26	4.28	1.00	22.69	7.68	**2.95**	64.87	14.90	**4.35**	28.67	9.81	**2.92**	-0.14	0.11	-1.22	0.58	0.19	**3.11**
	Dissimilar vs. Alternative	-4.54	4.16	-1.09	-5.09	7.78	-0.65	-5.76	16.57	-0.35	7.04	9.70	0.73	-0.04	0.13	-0.33	0.15	0.14	1.07
Sentence frame	High vs. Low	11.20	3.27	**3.43**	26.50	8.16	**3.25**	33.63	16.05	**2.10**	28.32	10.45	**2.71**	-0.05	0.14	-0.37	0.12	0.12	0.98
Interactions	Alternative vs. Identical × Sentence frame	-6.91	8.06	-0.86	6.52	14.05	0.46	-5.71	27.12	-0.21	-2.75	18.41	-0.15	0.01	0.22	0.05	-0.23	0.32	-0.72
	Dissimilar vs. Alternative × Sentence frame	6.23	7.92	0.79	-2.17	13.80	-0.16	-20.04	26.89	-0.75	-0.11	18.05	-0.01	0.46	0.23	**2.02**	-0.16	0.28	-0.57

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with |*t*| or |*z*| >= 1.96 are marked in bold.

Table 3c: LMM analyses on the post-target word by sentence frame. Each column represents a model.

	Sentence frames with high-frequency target words only	Skipping probability		
		b	**SE**	***z***
	(Intercept)	-0.91	0.11	**-8.11**
Preview	Alternative vs. Identical	-0.11	0.16	-0.72
	Dissimilar vs. Alternative	-0.33	0.18	-1.86

	Sentence frames with low-frequency target words only	Skipping probability		
		b	**SE**	***z***
	(Intercept)	-0.94	0.14	**-6.94**
Preview	Alternative vs. Identical	-0.13	0.17	-0.77
	Dissimilar vs. Alternative	0.15	0.17	0.90

b: Regression coefficient, SE: standard error, *t* or *z*: test statistic (*b*/SE). Cells with |*t*| or |*z*| >= 1.96 are marked in bold.

Figure captions

1. Example stimuli from the present study. While readers were fixating to the left of the invisible boundary (dashed lined), the target word in each of the sentence frames (dog/dim) was replaced with a preview as shown in the left column. After a reader crossed the boundary, the preview was replaced with the actual target word (right column).

Figure 1.

Sentence frame	Preview	Example sentence before display change	Example sentence after display change
High-frequency target	High-frequency	The excitable dog was eager to go for his walk.	The excitable dog was eager to go for his walk.
	Low-frequency	The excitable dim was eager to go for his walk.	
	Random letters	The excitable hev was eager to go for his walk.	
Low-frequency target	High-frequency	The increasingly dog light made it hard to see.	The increasingly dim light made it hard to see.
	Low-frequency	The increasingly dim light made it hard to see.	
	Random letters	The increasingly bup light made it hard to see.	

Appendix

Table A1: List of sentences used in the experiment along with predictability, violation, and target and preview word class criteria that were used for the supplementary analyses. Parentheses mark the target word and the alternative preview word. Note that each subject only saw one of the two sentences in a pair.

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
1	1	Last year, he lost (all/ant) his money on the stock market.	0.13	syntactic	closed	open
1	2	There was a massive (ant/all) infestation in the old house.	0.00	syntactic	open	closed
2	3	Today you must (buy/hag) the cake for the birthday party.	0.00	syntactic	open	open
2	4	In his dream, he saw an ancient (hag/buy) who scared him.	0.00	syntactic	open	open
3	5	My brother has terrible (aim/ape) and he will never play baseball.	0.00	syntactic	open	open
3	6	This weekend, I saw a large (ape/aim) and her new-born baby at the zoo.	0.00	semantic	open	open
4	7	Every Sunday, the little (boy/beg) would go to church with his family.	0.36	syntactic	open	open
4	8	The children would always (beg/boy) their mother for ice cream.	0.00	syntactic	open	open
5	9	She tried with all her might (but/mat) the heavy bookshelf would not budge.	0.27	syntactic	closed	open
5	10	The karate teacher threw his students down on the large (mat/but) multiple times.	0.91	syntactic	open	closed
6	11	Using the new system, the students (can/cow) access their grades online.	0.09	syntactic	closed	open
6	12	Early in the morning, he walked his only (cow/can) out to pasture.	0.00	semantic	open	open
7	13	It was an emotional (day/dew) for all of the family members.	0.27	semantic	open	open
7	14	They noticed the moist (dew/day) on the rose's petals.	0.09	semantic	open	open
8	15	They attempted to keep their feet (dry/duo) while crossing the small stream.	0.27	syntactic	open	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
8	16	The famous singing (duo/dry) performed for the entranced audience.	0.00	syntactic	open	open
9	17	The manager thought it would be great if everyone (did/don) more work in the office.	0.00	syntactic	open	open
9	18	He would never (don/did) the red hat for the formal ceremony.	0.00	syntactic	open	open
10	19	The excitable (dog/dim) was anxious to go for his walk.	0.18	syntactic	open	open
10	20	The increasingly (dim/dog) light made it hard to see.	0.00	syntactic	open	open
11	21	The liked the look of the large (cup/nip) that was for sale in the store.	0.00	semantic	open	open
11	22	My cat will sometimes (nip/cup) my hand when it is playful.	0.00	semantic	open	open
12	23	Jane rinsed her left (eye/emu) because something was in it.	0.18	semantic	open	open
12	24	With ruffled feathers, the angry (emu/eye) raced up and down his paddock.	0.00	semantic	open	open
13	25	The house was very (far/cap) from his school.	0.00	syntactic	open	open
13	26	His dirty (cap/far) did not make John look any better.	0.00	syntactic	open	open
14	27	After the long drive, we had to stop (for/rub) gas and supplies.	0.36	syntactic	closed	open
14	28	Her friend gave Maria a back (rub/for) and she felt much better.	0.27	syntactic	open	closed
15	29	The ranger reported seeing (few/fin) birds this year.	0.00	syntactic	closed	open
15	30	The fish was left with only a single (fin/few) after the attack.	0.09	semantic	open	closed
16	31	We filled our tank with (gas/shy) and then drove off into the night.	0.18	syntactic	open	open
16	32	The boy was very (shy/gas) when he performed on stage.	0.00	syntactic	open	open
17	33	She requested that I immediately (get/gym) her a chocolate bar from the store.	0.00	syntactic	open	open
17	34	I promised myself that I would go to the local (gym/get) every day this week.	0.00	syntactic	open	open
18	35	The man asked the tribal (god/gag) for relief from the persistent rain.	0.00	semantic	open	open
18	36	The burglar will (gag/god) the frightened old lady.	0.00	syntactic	open	open
19	37	She told her husband that they (had/oar) to pay the bills by the end of the month.	0.00	syntactic	closed	open
19	38	The man picked up the extra (oar/had) and prepared to row across the lake.	0.00	syntactic	open	closed
20	39	The old man across the street (has/hem) a very bad cold.	0.00	syntactic	closed	open
20	40	My grandmother told me she would (hem/has) that dress for me.	0.00	syntactic	open	closed
21	41	That night, she asked (her/hit) father if she could go to the party.	0.09	syntactic	open	open
21	42	The young child (hit/her) the ball over the fence.	0.00	syntactic	open	open
22	43	The woman asked (him/bug) if he knew where the High Street was.	0.09	syntactic	closed	open
22	44	During our road trip a huge (bug/him) splattered onto our window.	0.00	syntactic	open	closed

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
23	45	The room was incredibly (hot/hen) and the men decided to leave.	0.09	syntactic	open	open
23	46	The mother (hen/hot) watched over her chicks in the coop.	0.09	syntactic	open	open
24	47	The girl knew (how/hid) the magician's trick worked.	0.00	syntactic	closed	open
24	48	The little girl (hid/how) where no one could find her.	0.00	syntactic	open	closed
25	49	The puppy devoured (its/lip) dinner in a few short minutes.	0.00	semantic	closed	open
25	50	Mary's expensive (lip/its) treatment turned out to have no effect at all.	0.00	syntactic	open	closed
26	51	The office workers regularly (lay/rum) down the papers in just the right order.	0.00	semantic	open	open
26	52	Jacob enjoyed having (rum/lay) with his coffee.	0.00	semantic	open	open
27	53	I have never (met/pad) a celebrity before.	0.09	syntactic	open	open
27	54	She liked standing on the soft (pad/met) in the bedroom.	0.00	syntactic	open	open
28	55	Every year they travel somewhere (new/nap) like Greece or Brazil.	0.45	syntactic	open	open
28	56	My dog loves taking a long (nap/new) when its hot outside.	0.18	semantic	open	open
29	57	My baby cousin will not eat spinach (nor/net) broccoli but enjoys eating pizza.	0.00	syntactic	closed	open
29	58	The old man pulled in his heavy (net/nor) and discovered that he had caught a swordfish.	0.09	syntactic	open	closed
30	59	They want to leave (now/hog) rather than later in the evening.	0.09	syntactic	open	open
30	60	The husband would (hog/now) all of the blankets at night.	0.00	semantic	open	open
31	61	The volunteers wiped the toxic (oil/ode) off of the animals as best they could.	0.00	semantic	open	open
31	62	The children wrote a beautiful (ode/oil) to their teachers.	0.00	semantic	open	open
32	63	The manuscript was very (old/owl) and the team had to handle it carefully.	0.27	syntactic	open	open
32	64	I saw a large (owl/old) flying over the barn.	0.00	semantic	open	open
33	65	There was only (one/rid) doll to play with.	0.55	syntactic	open	open
33	66	He could finally (rid/one) himself of all the old paperwork.	0.00	syntactic	open	open
34	67	We walked out onto (our/orb) front porch with a cup of coffee.	0.00	syntactic	closed	open
34	68	A bright shimmering (orb/our) appeared in the night sky.	0.00	syntactic	open	closed
35	69	She could probably (pay/pea) for the couch with her credit card.	0.00	syntactic	open	open
35	70	She would not eat even a single (pea/pay) off her plate.	0.00	semantic	open	open
36	71	They had to pay fifty pounds (per/nod) person for tickets to the game.	0.00	syntactic	closed	open
36	72	The nice man would kindly (nod/per) every time someone entered the store.	0.00	syntactic	open	closed
37	73	Each student must (put/pew) their chairs on their desk at the end of the day.	0.00	syntactic	open	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
37	74	The caretaker bought an expensive (pew/put) for the church.	0.00	syntactic	open	open
38	75	The couple (own/sap) a large house in the country.	0.00	semantic	open	open
38	76	The sweet (sap/own) of some trees can be used for cooking.	0.00	semantic	open	open
39	77	There were only (six/rot) men in the hunting party.	0.00	syntactic	closed	open
39	78	The vegetables began to immediately (rot/six) after she had bought them.	0.18	syntactic	open	closed
40	79	They started the long and bloody (war/spy) over a silly disagreement.	0.18	semantic	open	open
40	80	The man would constantly (spy/war) on his neighbors.	0.00	semantic	open	open
41	81	Maybe today (you/rug) should clean the living room.	0.09	syntactic	closed	open
41	82	I spilled juice on the precious (rug/you) when she bumped my arm.	0.18	semantic	open	closed
42	83	The best team will (win/soy) the championship.	0.27	syntactic	open	open
42	84	She drinks (soy/win) milk only because she is lactose intolerant.	0.00	semantic	open	open
43	85	Jonathan walked (two/tug) miles yesterday.	0.00	syntactic	closed	open
43	86	She had to forcefully (tug/two) the shirt to get it out from under the dresser.	0.00	syntactic	open	closed
44	87	The little girl asked (why/wig) the sky is blue.	0.09	syntactic	closed	open
44	88	I wore my purple (wig/why) the other day for the fancy dress party.	0.00	syntactic	open	closed
45	89	Jane stated (who/wad) had robbed her.	0.00	syntactic	closed	open
45	90	She had a giant (wad/who) of notes in her wallet.	0.09	syntactic	open	closed
46	91	Despite the bad weather, John (was/tee) still planning to go to the concert.	0.18	syntactic	closed	open
46	92	She placed the ball on the small (tee/was) and hit it into the distance.	0.00	syntactic	open	closed
47	93	I asked her to send me the package (via/tub) air mail as soon as possible.	0.00	syntactic	closed	open
47	94	I like to relax in my large (tub/via) at the end of a long day.	0.00	syntactic	open	closed
48	95	To her dismay, Susan still cannot (use/web) the heavy weights at the gym.	0.00	syntactic	open	open
48	96	After her intricate (web/use) was destroyed, the spider was very confused.	0.00	syntactic	open	open
49	97	I like to make (tea/den) on hot summer days.	0.00	semantic	open	open
49	98	I saw the little foxes near the hidden (den/tea) when I was walking in the forest.	0.36	semantic	open	open
50	99	During dinner my father (sat/tip) at the head of the table.	0.00	syntactic	open	open
50	100	I left a large (tip/sat) for the helpful waiter.	0.09	syntactic	open	open
51	101	I was annoyed by the hefty (tax/urn) I had to pay.	0.00	semantic	open	open
51	102	They noticed the small (urn/tax) in the funeral director's office.	0.00	semantic	open	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
52	103	She asked me if there (are/pup) any doughnuts left over.	0.00	syntactic	closed	open
52	104	The campers stayed away from the small (pup/are) because they knew its mother was near.	0.00	syntactic	open	closed
53	105	Air traffic controllers must sometimes (act/lab) quickly in order to avoid a tragedy.	0.00	semantic	open	open
53	106	The student worked in the secret (lab/act) every weekend.	0.18	semantic	open	open
54	107	The man provided (aid/awe) to the woman before it was too late.	0.00	semantic	open	open
54	108	The critics expressed (awe/aid) at the orchestra's inspiring rendition of Beethoven's fifth.	0.00	semantic	open	open
55	109	The afternoon (sun/wry) was still very hot.	0.09	semantic	open	open
55	110	The audience liked the presenters and their (wry/sun) sense of humour.	0.00	semantic	open	open
56	111	To enter the mansion we needed to find the long (key/cub) hidden under a stone.	0.45	semantic	open	open
56	112	The scared (cub/key) stayed close to the mother bear.	0.00	semantic	open	open
57	113	The sum was so hard she couldn't even (try/fad) to solve it.	0.00	syntactic	open	open
57	114	The popular girl at school started a silly (fad/try) by wearing her socks inside-out.	0.09	semantic	open	open
58	115	The aggressive (cat/woo) refused to be touched by anyone.	0.00	semantic	open	open
58	116	The lobbyist will (woo/cat) politicians by giving them expensive gifts.	0.00	semantic	open	open
59	117	The cowboy used his loud (gun/aft) to scare the Indians.	0.09	syntactic	open	open
59	118	The captain told the trainee to stand at the very (aft/gun) of the boat.	0.00	syntactic	open	open
60	119	The man cheered when his favourite (son/coy) scored a goal.	0.00	syntactic	open	open
60	120	The woman was very (coy/son) with the younger gentleman.	0.00	syntactic	open	open
61	121	He stayed in the warm (sea/ark) until his fingers wrinkled.	0.00	semantic	open	open
61	122	He loaded up the wooden (ark/sea) with all of the animals.	0.00	Semantic	open	open
62	123	She told a very (bad/doe) lie to her teacher.	0.09	Syntactic	open	open
62	124	He saw the graceful (doe/bad) leap through the field.	0.00	Semantic	open	open
63	125	A few years (ago/bud) we traveled to Europe.	0.55	Syntactic	closed	open
63	126	Watching the flowers (bud/ago) was magical.	0.00	Syntactic	open	closed
64	127	She did not like having such a small (bed/tan) but she had no choice.	0.00	Semantic	open	open
64	128	They lay in the garden to gently (tan/bed) themselves in the sun.	0.00	Semantic	open	open
65	129	My birthday present came in a huge (box/din) that couldn’t even fit through my door.	0.64	Semantic	open	open
65	130	Earl couldn't endure the awful (din/box) for too long.	0.00	Semantic	open	open
66	131	It was during this (era/elf) that Christopher Columbus found the New World.	0.00	Semantic	open	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
66	132	At the front of the line, a happy little (elf/era) was waiting to greet the children.	0.09	Semantic	open	open
67	133	The officer asked if by any chance (she/sum) had seen anything.	0.00	Syntactic	closed	open
67	134	He could tell nobody about his terrible (sin/she) for many years.	0.00	Syntactic	open	closed
68	135	The man was told that he should (sit/toy) down and wait.	0.00	Syntactic	open	open
68	136	The child tried to steal the expensive (toy/sit) but he was caught.	0.00	Syntactic	open	open
69	137	In the forest, the giant (bee/tar) flew right past my head.	0.00	Semantic	open	open
69	138	Eric got his shoes stuck in the thick (tar/bee) and had trouble getting them free.	0.00	Semantic	open	open
70	139	For my sister's birthday she asked for an apple (pie/pro) instead of a cake.	0.36	Semantic	open	open
70	140	They took lessons from the tanned (pro/pie) at the tennis club.	0.00	Semantic	open	open
71	141	I told my friend (not/nub) to disturb my father while he was working.	0.00	Syntactic	closed	open
71	142	She sharpened her pencil down to a tiny (nub/not) in the middle of class.	0.09	Syntactic	open	closed
72	143	Arthur's sore (leg/pat) caused him a lot of pain.	0.00	Semantic	open	open
72	144	He gave his colleague a friendly (pat/leg) on the back.	0.36	Semantic	open	open
73	145	He asked to be immediately (let/mop) into the house, since it was raining so terribly.	0.00	Syntactic	open	open
73	146	My dad told me to use our heavy (mop/let) in order to clean up the spilled juice.	0.00	Syntactic	open	open
74	147	The man filled his small (bag/fog) with only the things he needed.	0.27	Semantic	open	open
74	148	The bus driver was driving slow because heavy (fog/bag) limited his visibility.	0.00	Semantic	open	open
75	149	They told him to sell (his/lap) car this year and use a bike instead.	0.27	Syntactic	closed	open
75	150	He collapsed on the last (lap/his) of the race, just a few feet from the finish line.	0.09	Syntactic	open	closed
76	151	I wish I could (fly/gin) so I could see the world from above.	0.00	Syntactic	open	open
76	152	The old man ordered (gin/fly) at the hotel bar.	0.00	Semantic	open	open
77	153	The strange (guy/icy) in Anna's theatre class makes her feel uncomfortable.	0.00	Semantic	open	open
77	154	The road was extremely (icy/guy) and the driver could not continue.	0.00	Syntactic	open	open
78	155	The specialist tried to see as many (ill/rip) people as possible.	0.00	Syntactic	open	open
78	156	It is easy to accidentally (rip/ill) this dress.	0.00	Syntactic	open	open
79	157	She really wanted to wear the blue (top/wit) that was on sale.	0.09	Semantic	open	open
79	158	His biting (wit/top) won the comedian many admirers.	0.00	Semantic	open	open
80	159	It was not time to leave (yet/cot) and so the kids just played in the street.	0.45	Syntactic	closed	open
80	160	The child in the small (cot/yet) slept soundly.	0.00	Syntactic	open	closed

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
81	161	I will tell her I have (got/kit) the crisps and peanuts for the party.	0.00	Syntactic	closed	open
81	162	When he bought the chemistry (kit/got) it didn't come with instructions.	0.18	Syntactic	open	closed
82	163	The school boy would (run/pox) everyday so he could stay in shape.	0.09	Semantic	open	open
82	164	The expert was worried about the spread of the dangerous (pox/run) virus this year.	0.00	Semantic	open	open
83	165	The boy felt he had become a real (man/sax) after his hard Summer work on the farm.	0.09	Semantic	open	open
83	166	The magnificent (sax/man) solo was what jazz musician was famous for.	0.00	Semantic	open	open
84	167	The purple (pen/hop) spilt all over my shirt.	0.00	Semantic	open	open
84	168	I trained my pet so it will (hop/pen) on command.	0.00	Semantic	open	open
85	169	I couldn't think of a good (end/ray) to the story, so I needed help.	0.00	Semantic	open	open
85	170	She directed a strong (ray/end) of light onto the wall using a prism.	0.00	Semantic	open	open
86	171	He broke the local (law/dye) and he's going to jail.	0.00	Semantic	open	open
86	172	She could (dye/law) her hair if she wanted to look different.	0.00	Syntactic	open	open
87	173	They looked at the afternoon (sky/cut) and admired the clouds.	0.09	Semantic	open	open
87	174	She accidentally (cut/sky) herself while making a collage.	0.00	Syntactic	open	open
88	175	The skyscraper didn't look very (big/dab) from the airplane.	0.09	Syntactic	open	open
88	176	She added another (dab/big) of paint to her canvas, and then stopped for the day.	0.00	Semantic	open	open
89	177	The strong (men/fun) lifted the wheels.	0.00	Semantic	open	open
89	178	They knew that much (fun/men) was to be had at the carnival.	0.00	Syntactic	open	open
90	179	The class took a public (bus/err) to the museum.	0.27	Syntactic	open	open
90	180	Even the best mathematicians will sometimes (err/bus) in their calculations.	0.09	Semantic	open	open
91	181	The bird watcher (saw/ore) the most amazing blue jay yesterday.	0.00	Syntactic	open	open
91	182	The explorers were very excited because they found (ore/saw) on the small island.	0.00	Syntactic	open	open
92	183	The spy crawled very (low/bun) to the ground so the guards couldn't see him.	0.00	Syntactic	open	open
92	184	We only had a single (bun/low) and one tomato, so we couldn't make hamburgers.	0.00	Semantic	open	open
93	185	He could (see/foe) the entire valley with his binoculars.	0.00	Syntactic	open	open
93	186	After his cunning (foe/see) had been defeated, the hero was relieved.	0.00	Semantic	open	open
94	187	After spending a few minutes in the humid (air/lag) she collapsed.	0.36	Semantic	open	open
94	188	Many countries still (lag/air) behind others in terms of environmental awareness.	0.00	Semantic	open	open
95	189	We decided to stop (and/kid) help the stranded driver.	0.18	Syntactic	closed	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
95	190	The old woman told the little (kid/and) to stop yelling.	0.00	Semantic	open	closed
96	191	At the desk, they will (ask/elm) her if she packed her own luggage.	0.09	Syntactic	open	open
96	192	She planted a tiny (elm/ask) tree near the local library.	0.00	Syntactic	open	open
97	193	The captain made the guard (arm/inn) himself for battle.	0.00	Syntactic	open	open
97	194	They decided to stay at a lovely (inn/arm) for the night.	0.09	Semantic	open	open
98	195	In the military, it is customary to call one's superiors (sir/jog) as a sign of respect.	0.45	Semantic	open	open
98	196	The woman was tired from her long (jog/sir) through the park.	0.00	Semantic	open	open
99	197	During the flight Harriet (sat/dip) near the window so she could get a good view.	0.00	Syntactic	open	open
99	198	We went for a pleasant (dip/sat) after our long hot day.	0.00	Syntactic	open	open
100	199	He could not understand his brother's slightly (odd/par) behaviour these days.	0.00	Semantic	open	open
100	200	They all tried to remain under (par/odd) but didn't succeed.	0.00	Syntactic	open	open
101	201	The team were very (sad/sew) because they lost the championship.	0.00	Syntactic	open	open
101	202	Agatha's grandmother will always (sew/sad) in her free time.	0.00	Syntactic	open	open
102	203	The student (ran/oak) so fast his shoes almost fell off.	0.00	Syntactic	open	open
102	204	From the top of the tall (oak/ran) the campers could see how big the forest truly was.	0.00	Syntactic	open	open
103	205	My friends took me to the fancy (pub/rev) for my birthday.	0.00	Semantic	open	open
103	206	The teenager would (rev/pub) his engine at every stop-light.	0.00	Syntactic	open	open
104	207	The arena turned into (mud/jot) after the rain.	0.00	Syntactic	open	open
104	208	The student had to quickly (jot/mud) down the homework before the teacher erased it.	0.00	Syntactic	open	open
105	209	If the team stays disciplined they should at least (tie/tag) against their opponents.	0.00	Semantic	open	open
105	210	The boy's neck itched because of the uncomfortable (tag/tie) on his shirt.	0.00	Semantic	open	open
106	211	He found the shoes a tiny (bit/vow) too small.	0.82	Semantic	open	open
106	212	You must (vow/bit) to be with another person for the rest of your life when you get married.	0.00	Syntactic	open	open
107	213	Before she could (say/tin) anything, the salesman cut her off.	0.00	Semantic	open	open
107	214	They used (tin/say) to fix their toy boat.	0.00	Syntactic	open	open
108	215	The worker picked up his enormous (axe/pun) and cut down the large tree.	0.27	Semantic	open	open
108	216	The speaker made a very clever (pun/axe) during dinner.	0.09	Semantic	open	open
109	217	The fruit was very (red/tow) and looked delicious.	0.00	Syntactic	open	open
109	218	Tomorrow they will (tow/red) their caravan to the seaside.	0.00	Syntactic	open	open

Pair	Sentence Number	Sentence	Cloze Predictability	Violation	Target word class	Preview word class
110	219	The teacher was very (mad/lob) when she found out that the students had cheated.	0.09	Syntactic	open	open
110	220	The instructor will simply (lob/mad) the ball at the weaker students.	0.00	Syntactic	open	open
111	221	To get hired for the competitive (job/zip) the woman had to work sixty hours a week.	0.36	Semantic	open	open
111	222	The cars will (zip/job) past the observers during the race.	0.00	Syntactic	open	open
112	223	The girl asked her parents if she could play outside (too/hue) and they consented.	0.00	Syntactic	closed	open
112	224	The curator pointed out the beautiful (hue/too) of the photo.	0.00	Syntactic	open	closed
113	225	The child always asks very politely if he and his brother (may/zoo) go outside.	0.00	Syntactic	closed	open
113	226	Our school went on a field trip to the renowned (zoo/may) last Friday.	0.00	Syntactic	open	closed
114	227	There is now a strict (ban/wed) on smoking in restaurants.	0.00	Syntactic	open	open
114	228	The couple was finally (wed/ban) last autumn.	0.00	Syntactic	open	open
115	229	My friend misread the complicated (map/hug) so we got lost.	0.00	Semantic	open	open
115	230	She gave a tight (hug/map) to her Mum when she saw her.	0.36	Semantic	open	open
116	231	After my brother picked the wrong (way/sly) we got lost for hours.	0.00	Syntactic	open	open
116	232	The exceedingly (sly/way) thief had no problems stealing the painting.	0.00	Syntactic	open	open
117	233	Her friends took her to a small (bar/paw) for her birthday.	0.00	Semantic	open	open
117	234	The injured (paw/bar) had to be cleaned and bandaged.	0.00	Semantic	open	open
118	235	She wouldn't (fit/flu) into her old prom dress if she tried it on now.	0.00	Syntactic	open	open
118	236	The man thought he had severe (flu/fit) but he really just had a cough.	0.00	Syntactic	open	open
119	237	The child did not know the numbers after (ten/yew) yet but learned quickly.	0.45	Syntactic	closed	open
119	238	There was a tall (yew/ten) tree behind the house.	0.00	Semantic	open	closed
120	239	The dogs were extremely (wet/wag) so we didn't let them in the house.	0.00	Syntactic	open	open
120	240	With a quick (wag/wet) of its tail the dog bounded in the house.	0.00	Syntactic	open	open

Table A2: Word frequency and mean letter bigram (token) frequency for each target word (taken from the CELEX database using the *N-Watch* software, Davis, 2005).

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
1	1	all	3597.49	3630.37
1	2	ant	3.85	15047.34
2	3	buy	126.2	2900.31
2	4	hag	1.01	4274.34
3	5	aim	40.89	1469.16
3	6	ape	5.08	15.27
4	7	boy	216.54	324.75
4	8	beg	11.62	209.49
5	9	but	5412.79	7233.24
5	10	mat	7.54	1406.73
6	11	can	2081.84	2908.21
6	12	cow	22.51	1661.13
7	13	day	766.98	2573.04
7	14	dew	5.2	836.78
8	15	dry	89.55	256.79
8	16	duo	0.67	57.24
9	17	did	1170.11	1321.28
9	18	don	79.22	245.05
10	19	dog	71.73	134.3
10	20	dim	15.7	1942.18
11	21	cup	60.84	155.45
11	22	nip	1.96	42.47
12	23	eye	127.6	144.25
12	24	emu	N/A	N/A
13	25	far	515.64	942.2
13	26	cap	30.34	1295.7

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
14	27	for	8288.04	8394.62
14	28	rub	12.74	160.43
15	29	few	585.08	1150.69
15	30	fin	3.63	132.26
16	31	gas	70.89	6587.67
16	32	shy	18.04	2399.6
17	33	get	1169.94	1945.94
17	34	gym	4.08	4.08
18	35	god	260.78	719.91
18	36	gag	1.84	105.56
19	37	had	6255.03	7541.48
19	38	oar	0.39	640.57
20	39	has	2123.13	10756.46
20	40	hem	3.46	1947.06
21	41	her	3854.97	4054.32
21	42	hit	91.34	4402.21
22	43	him	2515.31	5395.71
22	44	bug	3.24	2837.49
23	45	hot	139.55	3909.55
23	46	hen	5.59	2431.92
24	47	how	1183.8	2286.02
24	48	hid	11.96	4774.81
25	49	its	1552.12	1552.12
25	50	lip	16.98	116.62
26	51	lay	156.15	2351.93
26	52	rum	5.53	167.97
27	53	met	148.16	1760.23
27	54	pad	11.79	3442.71

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
28	55	new	1062.07	1377.48
28	56	nap	5.03	78.8
29	57	nor	171.96	7779.9
29	58	net	32.35	1905.9
30	59	now	1802.12	5167.04
30	60	hog	2.4	718.96
31	61	oil	123.85	128.99
31	62	ode	1.01	30.87
32	63	old	752.35	752.35
32	64	owl	3.02	467.29
33	65	one	3455.7	3455.7
33	66	rid	35.36	699.83
34	67	our	1286.54	2562.32
34	68	orb	0.45	1.99
35	69	pay	189.66	2300.55
35	70	pea	1.68	333.85
36	71	per	363.97	2314.75
36	72	nod	11.12	3706.87
37	73	put	687.26	4780.78
37	74	pew	2.4	1034.94
38	75	own	916.03	923.8
38	76	sap	2.57	847.23
39	77	six	211.96	394.11
39	78	rot	8.32	3286.06
40	79	war	339.78	6841.37
40	80	spy	8.38	9.24
41	81	you	7189.27	7190.19
41	82	rug	11.68	163

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
42	83	win	62.35	122.96
42	84	soy	2.18	227.01
43	85	two	1371.62	1371.62
43	86	tug	5.42	35.6
44	87	why	620.28	1832.43
44	88	wig	6.7	226.43
45	89	who	2395.59	2705.73
45	90	wad	3.24	9519.41
46	91	was	10857.38	12734.38
46	92	tee	4.08	779.91
47	93	via	19.61	21.17
47	94	tub	7.88	33.03
48	95	use	485.98	485.98
48	96	web	5.81	42.49
49	97	tea	88.77	295.11
49	98	den	9.05	494.19
50	99	sat	228.04	1071.58
50	100	tip	24.36	85.03
51	101	tax	108.88	134.1
51	102	urn	2.63	2.63
52	103	are	4424.36	4577.31
52	104	pup	0.89	387.85
53	105	act	187.15	189.02
53	106	lab	9.94	196.03
54	107	aid	56.76	848.26
54	108	awe	8.27	15.14
55	109	sun	153.3	352.83
55	110	wry	2.63	213.33

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
56	111	key	71.56	89.43
56	112	cub	2.57	150.07
57	113	try	268.38	346.21
57	114	fad	3.13	3620.25
58	115	cat	41.28	1520.04
58	116	woo	2.4	573.53
59	117	gun	63.58	319.73
59	118	aft	1.06	1.62
60	119	son	159.66	251.92
60	120	coy	1.84	179.9
61	121	sea	160.45	1045.89
61	122	ark	2.85	2358.74
62	123	bad	209.78	3510.2
62	124	doe	2.74	98.94
63	125	ago	225.14	354.47
63	126	bud	3.85	2826.43
64	127	bed	244.47	483.41
64	128	tan	7.37	1761.53
65	129	box	78.66	236.15
65	130	din	4.47	733.32
66	131	era	23.24	31.84
66	132	elf	0.34	4.33
67	133	she	4132.18	34687.18
67	134	sin	25.31	349.21
68	135	sit	119.94	564.58
68	136	toy	14.58	842.51
69	137	bee	6.59	781.89
69	138	tar	4.64	707.48

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
70	139	pie	12.57	138.43
70	140	pro	5.81	10.26
71	141	not	5109.72	6790.56
71	142	nub	0.73	32.49
72	143	leg	63.52	348.64
72	144	pat	19.05	430.54
73	145	let	393.3	1665.73
73	146	mop	4.08	150.52
74	147	bag	62.63	243.06
74	148	fog	9.39	4214.7
75	149	his	5576.54	6894.42
75	150	lap	18.66	257.58
76	151	fly	50.95	56.79
76	152	gin	15.25	90.03
77	153	guy	56.48	157.94
77	154	icy	9.11	35.2
78	155	ill	62.63	1861.66
78	156	rip	4.58	69.25
79	157	top	236.7	843.65
79	158	wit	12.63	338.32
80	159	yet	469.22	1982.37
80	160	cot	10.73	3284.65
81	161	got	860.11	3803.6
81	162	kit	8.49	319.69
82	163	run	229.89	387.61
82	164	pox	1.51	74.1
83	165	man	1061.79	2794.89
83	166	sax	N/A	832.31

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
84	167	pen	19.44	692.34
84	168	hop	5.08	805.94
85	169	end	496.7	14880.67
85	170	ray	12.91	2263.97
86	171	law	166.65	488.64
86	172	dye	5.7	83.3
87	173	sky	77.09	80.22
87	174	cut	177.88	4548.38
88	175	big	317.15	475.47
88	176	dab	2.4	417.14
89	177	men	656.15	889.22
89	178	fun	45.98	286.18
90	179	bus	64.53	2842.29
90	180	err	1.4	18.47
91	181	saw	387.88	1078.3
91	182	ore	3.07	2221.76
92	183	low	144.58	1860.36
92	184	bun	3.74	3062.1
93	185	see	1171.28	1530.69
93	186	foe	13.91	4179.33
94	187	air	251.17	388.45
94	188	lag	3.3	226.93
95	189	and	28767.93	29677.73
95	190	kid	32.12	692.29
96	191	ask	226.42	249.38
96	192	elm	6.7	7.51
97	193	arm	106.03	2410.33
97	194	inn	9.44	14.3

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
98	195	sir	167.09	477.37
98	196	jog	3.35	197.49
99	197	sat	228.04	1071.58
99	198	dip	8.66	690.7
100	199	odd	59.72	102.29
100	200	par	5.25	764.66
101	201	sad	46.2	4083.74
101	202	sew	2.51	1746.97
102	203	ran	111.23	1782.13
102	204	oak	14.19	15.2
103	205	pub	20.73	382.46
103	206	rev	6.48	103.85
104	207	mud	29.22	42.46
104	208	jot	3.13	3388.07
105	209	tie	35.47	154.66
105	210	tag	5.7	118.38
106	211	bit	240.67	587.37
106	212	vow	5.2	1622.15
107	213	say	878.27	2941.59
107	214	tin	28.66	127.65
108	215	axe	5.64	5.64
108	216	pun	1.34	609.64
109	217	red	188.94	415.51
109	218	tow	3.41	2323.73
110	219	mad	48.21	4418.89
110	220	lob	1.45	395.7
111	221	job	244.25	299.89
111	222	zip	2.35	40.67

Pair	Sentence Number	Target Word	CELEX frequency	Bigram token frequency
112	223	too	1043.91	1239.14
112	224	hue	4.58	85.19
113	225	may	1057.32	3276.73
113	226	zoo	8.72	538.75
114	227	ban	13.8	1886.2
114	228	wed	3.3	354.47
115	229	map	32.18	1182.38
115	230	hug	4.64	53.44
116	231	way	1205.03	8377.26
116	232	sly	5.7	31.99
117	233	bar	67.88	832.16
117	234	paw	3.18	437.27
118	235	fit	69.94	347.62
118	236	flu	4.36	29.84
119	237	ten	226.48	653.6
119	238	yew	1.56	1453.95
120	239	wet	62.51	1397.65
120	240	wag	1.34	6252.27